

## Original Article

# Validation and public health modelling of risk prediction models for kidney cancer using the UK Biobank

Hannah Harrison<sup>1</sup> , Lisa Pennells<sup>1</sup>, Angela Wood<sup>1</sup>, Sabrina H. Rossi<sup>2</sup>, Grant D. Stewart<sup>2</sup> , Simon J. Griffin<sup>1</sup> and Juliet A. Usher-Smith<sup>1</sup> 

<sup>1</sup>Department of Public Health and Primary Care, University of Cambridge, and <sup>2</sup>Department of Surgery, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

## Objectives

To externally validate risk models for the detection of kidney cancer, as early detection of kidney cancer improves survival and stratifying the population using risk models could enable an individually tailored screening programme.

## Methods

We validated the performance of 30 existing phenotypic models predicting the risk of kidney cancer in the UK Biobank cohort ( $n = 450\,687$ ). We compared the discrimination and calibration of models for men, women, and a mixed-sex cohort. Population level data were used to estimate model performance in a screening scenario for a range of risk thresholds (6-year risk: 0.1–1.0%).

## Results

In all, 10 models had reasonable discrimination (area under the receiver-operating characteristic curve  $>0.60$ ), although some had poor calibration. Modelling demonstrated similar performance of the best models over a range of thresholds. The models showed an improvement in ability to identify cases compared to age- and sex-based screening. All the models performed less well in women than men.

## Conclusions

The present study is the first comprehensive external validation of risk models for kidney cancer. The best-performing models are better at identifying individuals at high risk of kidney cancer than age and sex alone; however, the benefits are relatively small. Feasibility studies are required to determine applicability to a screening programme.

## Keywords

kidney cancer, public health modelling, risk models, risk stratification, screening, validation, #kcs, #KidneyCancer, #uroonc

## Introduction

Kidney cancer is the seventh most common cancer in Europe [1], responsible for 50 000 annual deaths [2,3]. Incidence is projected to rise in coming years, due to changes in population demographics and lifestyle [1]. Early-stage diagnosis is strongly correlated with improved survival rates; The 5-year cancer-specific survival rates for patients diagnosed with Stage I and IV kidney cancer are 83% and 6%, respectively [4]; in the UK ~25% of patients who present with kidney cancer have evidence of metastases. In the UK, 60% of kidney cancers are diagnosed incidentally with ~20% of these being late stage (III–IV) [5]. A lack of symptoms,

even at late stages of the disease, makes the early detection of kidney cancer particularly challenging.

There are several risk factors that are known to be associated with kidney cancer and these have been combined to develop a number of risk models, as identified in a recent systematic review [6]. It has been suggested that such risk models could be used to identify high-risk individuals for inclusion in a screening programme [7], with the development and validation of risk models being identified as a research priority in a recent review of screening for kidney cancer [8]. Risk-stratified screening could improve screening efficiency, while reducing the associated costs and harms compared to a

simple age-based approach [9,10]; this is crucial when considering diseases, such as kidney cancer, with a relatively low prevalence [11]. However, the performance of most of the published models that predict individual-level risk of developing kidney cancer have not been externally validated. Therefore, the potential of these models to assess eligibility is unclear and comparisons to other approaches (such as using age- and sex-based selection) is not possible.

In the present study, we externally validated 30 of these models in the UK Biobank (UKB), a cohort of >500 000 people. We then applied a public health modelling approach to estimate the population level health benefit of incorporating the best performing models within a risk-stratified kidney cancer screening programme. The use of risk models to determine eligibility for screening was compared to age- and sex-based strategies.

## Methods

### Selection of Risk Prediction Models

We identified risk prediction models from our recent systematic review [6]. We found 60 studies describing models that predict the risk of an individual developing kidney cancer. Only models with phenotypic risk factors were included in this validation. Where there was insufficient detail about the models, the authors were contacted; these models were excluded if additional information was not obtained ( $n = 8$ ). Models including more than one risk factor for which there is no comparable UKB variable were excluded ( $n = 8$ ). A total of 16 studies (29 models) were included in this validation study [7,12–27] (Fig. S1). Following discussion with experts, an additional study published after the systematic review was included [7]. Details of the development studies, including study type, population, sample size and location are given in Table S1.

### Validation Cohort

The UKB was used as the validation cohort. It is the largest population-based cohort in the UK [28]; between the years 2006 and 2010, ~500 000 individuals aged 40–69 years, were recruited. Demographic and lifestyle information was recorded at baseline assessment; however, no imaging was carried out. Details of how risk factors were used is provided in Table S5. Data on cancer incidence is available through linkage to national cancer registries (Methods S1).

Members of the cohort with a diagnosis of kidney cancer prior to baseline were excluded from analysis. If more than one diagnosis was recorded, the first occurrence was used. The most recent cancer diagnosis in the dataset is December 2016. We censored all follow-up to 31 March 2016 (the study end date) to ensure that late registrations were not missed. A closed cohort analysis was conducted; individuals whose

follow-up was censored before 6 years were excluded. Cases were those who developed kidney cancer within 6-years of baseline assessment. This follow-up time is longer than the estimated sojourn time of RCC (3.7–5.8 years [29,30]).

### Model Validation

For all included models, we first computed the score, relative risk, or absolute risk (depending on model type) for each eligible UKB participant at baseline. Models predicting the absolute risk over a specific period were re-scaled to predict 6-year risk (Methods S1). We calculated the performance of the models both separately for men and women and as a whole cohort. Models developed originally for single sex cohorts were not validated for the other sex or the whole cohort. Where a study developed separate models for men and women the sex-specific risk predictions were combined for validation in the whole cohort.

A complete case approach was used for the primary analysis with each model only computed for individuals with data for all the risk factors. This was performed on a model-by-model basis; hence, the cohort size varies slightly for each validation. The outcome (diagnosis of kidney cancer) was treated as a binary variable. The discriminative ability of each model was measured quantitatively using the area under the receiver-operating characteristic curve (AUROC). The calibration of the models was assessed graphically using deciles (or maximum number of possible groups) of predicted risk (Methods S1). The observed to expected ratio (O:E) and the Hosmer–Lemeshow statistic were calculated for absolute risk models.

Sensitivity analyses were carried out. An open cohort analysis included all participants with <6 years follow-up and cases of kidney cancer occurring after 6 years. In that analysis, we used Harrell's C-statistic to assess discrimination, as it accounts for censoring. Secondly, we performed a complete-case analysis using a subset of participants with complete data available for all models.

### Public Health Modelling

Further analyses were carried out to assess the potential benefits of using these models in the UK general population. Five models predicting absolute risk, with good performance in validation (AUROC >0.60,  $0.5 < O:E < 2$ ) were selected for this analysis [31].

First, we used a re-calibration method (previously described [32,33]) to re-scale the risk distributions of the models to reflect the expected risk distribution in the general population (Methods S1). Second, we developed a set of scenarios using different risk thresholds (0.1–1.0% 6-year risk) as a cut-off for screening eligibility. For each model, members of the cohort with a predicted risk higher than the threshold were classified

as high-risk and hence 'eligible' for screening. Cases of kidney cancer in the high-risk group are considered identifiable in that screening scenario.

The proportion of true positives, false positives, false negatives, and true negatives were calculated for each sex and 5-year age group. These were then summed with weighting that reflected the age and sex distribution and the incidence of kidney cancer in the UK (using Cancer Research UK [CRUK] and Office for National Statistics [ONS] data) [34]. Using this new representative population, the efficacy of the model was determined for each screening scenario. An overview of this analysis process is given in the Methods S1.

We assessed accuracy of the models by calculating the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for each threshold. The efficiency of each screening scenario was assessed by calculating the proportion of the population invited to screening, the proportion of people subsequently diagnosed with kidney cancer within current routine care during the following 6 years who are invited, and the number of people needed to invite to identify one individual who is subsequently diagnosed with kidney cancer within 6 years (NNI). These screening scenarios were compared to age- and sex-based risk stratification. Sub-group analyses by sex and age group were carried out for one threshold (0.4%), selected pragmatically to represent a potential cut-off that selects ~10% of the population. This modelling approach assesses the potential of the risk models to determine screening eligibility; however, the practicalities of screening (such as the performance of screening tests and attendance at screening) are not factored into this analysis and the results are not dependent on screening modality.

## Results

A cohort of 450 687 individuals, including 635 cases of kidney cancer, were included in the primary analysis (Fig. 1). Compared to those who did not develop kidney cancer, the cases were more likely to be male, White, overweight, an ever smoker (and have higher pack-years), hypertensive, and have a previous cancer diagnosis (Table 1).

The models either assign scores ( $n = 4$ ) [13,19], calculate a relative risk ( $n = 17$ ) [12,14–18,20,22–24,26,35], or calculate an absolute risk over a defined period ( $n = 9$ ) [7,21,25,27] (Table 2, [7,12–27]). The most commonly included risk factors were body mass index (BMI;  $n = 29$ ), smoking ( $n = 22$ ), hypertension ( $n = 18$ ), and age ( $n = 13$ ). Five studies with mixed sex development populations published separate models for men and women [12,17,19,23,27]. A total of 10 models included sex as a distinct risk factor [7,20–22,25,26]. None of the variables used in the models had >2.1% missing data in the UKB cohort (Table S5).

## Discrimination

Figure 2A–C shows AUROC values for the models validated in men, women, and the whole cohort. Most of the models have poor discriminative ability;  $0.5 < \text{AUROC} < 0.6$ . The Singleton et al. [7], Scelo et al. [25], Usher-Smith et al. [27] and four of the Landsman and Graubard *al.* [21] models have reasonable discrimination in the mixed sex cohort ( $0.67 < \text{AUROC} < 0.71$ ), men ( $0.64 < \text{AUROC} < 0.71$ ) and women ( $0.62 < \text{AUROC} < 0.72$ ). Originally developed in women, the two Sanfilippo et al. [24] models have reasonable discriminative ability in women ( $0.61 < \text{AUROC} < 0.70$ ).

These 10 models all include age as a risk factor. Several also include blood pressure as a risk factor, either a hypertension diagnosis [7,21] or a direct measurement [7,24]. The results of the sensitivity analyses were consistent with the primary analysis (Table S2).

## Calibration

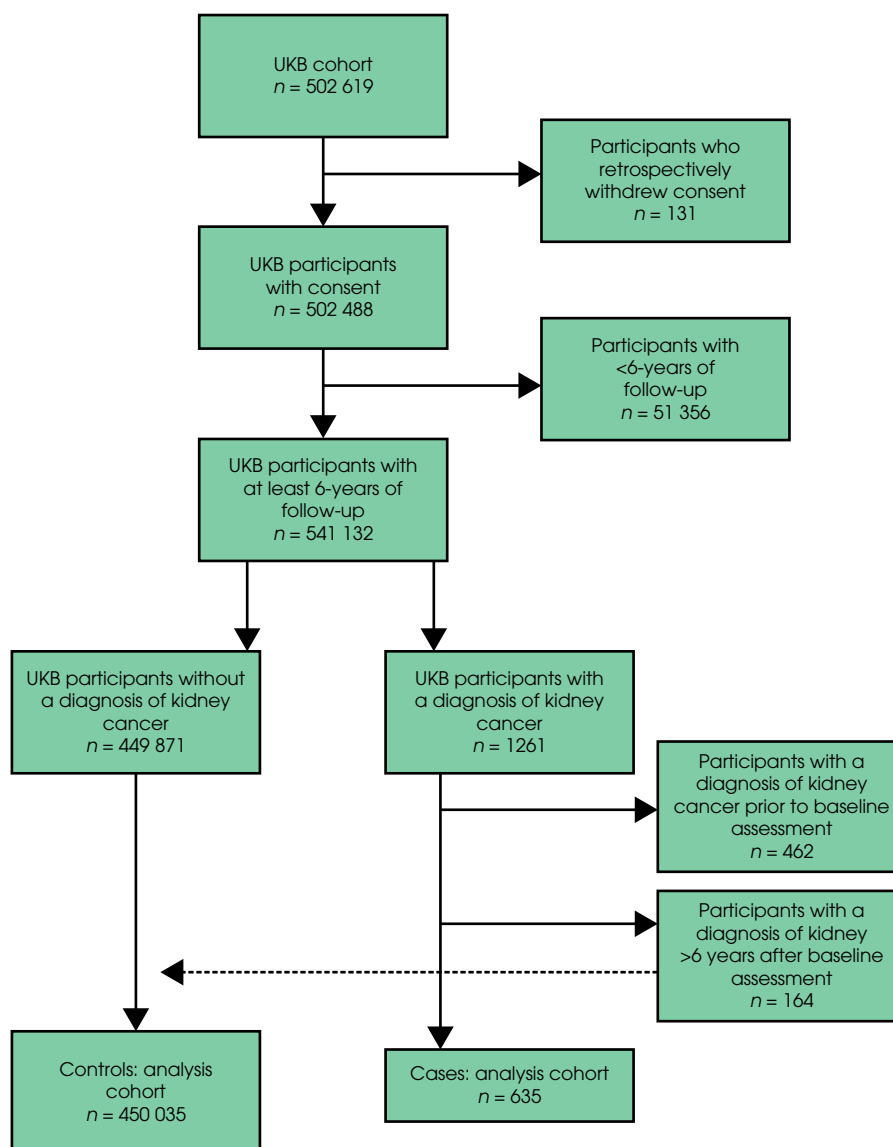
The calibration of the models in the UKB cohort is shown graphically in Fig. S2a–e for men, women, and the whole cohort. Some of the models show good calibration (Landsman and Graubard [21] 2013 [b]); however, most show discrepancies from ideal calibration. Common issues include overestimation of the risk in the lower deciles (Scelo et al. [25] 2018), good correlation but poor calibration (Landsman and Graubard [21] 2013 [d]) and no clear correlation (Landsman and Graubard [21] 2013 [a]).

Three of the absolute risk models (Singleton et al. [7] 2020, Landsman and Graubard [21] 2013 [b] and [c]) have non-significant Hosmer–Lemeshow tests, indicating well calibrated models (Table S3). In the mixed sex cohort, these three models have O:E values of 1.1, 1.1, and 1.2. The Scelo et al. [25] and Usher-Smith et al. [27] models have O:E values of 0.6 and 0.7, showing reasonable calibration.

## Public Health Modelling

The adjusted population used in this analysis was scaled to have 100 000 individuals aged 40–70 years. This population has the same age and sex distribution (within this age range) as the UK and the same incidence of kidney cancer. Over a 6-year period, 212 population members (139 men, 72 women) are expected to develop kidney cancer. If a screening programme included all 100 000 individuals and detected all cases (over 6-years), 471 individuals would be invited per case identified (NNI).

The Usher-Smith et al. [27], Scelo et al. [25], Singleton et al. [7], Landsman and Graubard [21] 2013 (b) and (c) models were included in this analysis. Table 3 [7,21,25,27] and Fig. 3 compare the performance of the models and other screening

**Fig. 1** Cohort selection process. This flow diagram describes the process used to select a cohort from UKB for the primary analysis in this study.

strategies over the range of risk thresholds (0.1–1.0%). For all models, as the risk threshold (and the proportion of the population screened) increases, the sensitivity of the models increases while the specificity (and NNI) decreases. Very few individuals were predicted to have a risk >1% for any model. No meaningful differences are seen between the five models.

At a risk threshold of 0.4% the NNI is <200 for all five models. The Landsman and Graubard [21] (b) model detects the highest number of kidney cancer cases, 32.5% (69/212), screening 12.7% of the population with a corresponding NNI of 183. The Scelo *et al.* [25] model detects the least number of cases at this threshold, 27.8% (59/212); however, a slightly

lower proportion of the population are screened (11.7%), with a corresponding NNI of 199. As seen in Fig. 3, the models performed slightly better than the age- and sex-based screening strategies. If all men aged ≥60 years in the modelling population were screened (14.1%), then 29.7% (63/212) of cases are identified (NNI = 206).

Table S4 compares the model performance at a risk threshold of 0.4% (a) to the age- and sex-based strategies (b) broken down by sex and 5-year age group. For all five models, the subgroup with the oldest men (65–69 years) has the highest screening rate and the highest detection rate; the two Landsman and Graubard [21] models have the lowest NNI (178) in this subgroup.

**Table 1** Summary of study cohort.

	All	Controls	Cases	P*	% Incident KCa	Odds ratio (95% CI)
Counts	450 670	450 035	635		0.141	
Age, years	Mean (SD) Missing, n (%)	56.5 (8.09) 0 (0)	61.0 (6.22) 0 (0)	<0.001	–	1.08 (1.07–1.10)
Sex	Female, n (%) Male, n (%) Missing, n (%)	245 029 (54.4) 205 641 (45.6) 0 (0)	227 (35.8) 408 (64.3) 0 (0)	– <0.001	0.0926 0.1984	1. (ref) 2.14 (1.82–2.52)
Ethnicity	White, n (%) Mixed heritage, n (%) Southeast Asian, n (%) Black, n (%) Chinese, n (%) Other, n (%) Missing, n (%)	426 076 (94.5) 2516 (0.56) 8265 (1.83) 6362 (1.41) 1347 (0.30) 3793 (0.84) 2311 (0.51)	425 454 (94.5) 2515 (0.56) 8259 (1.84) 6359 (1.41) 1347 (0.30) 3791 (0.84) 2310 (0.51)	0.021	0.1460 0.0397 0.0726 0.0472 0.0000 0.0527 0.0433	1. (ref) 0.27 (0.04–1.93) 0.50 (0.22–1.11) 0.32 (0.10–1.00) – 0.36 (0.90–1.45) –
BMI, kg/m <sup>2</sup>	Median (IQR) <20, n (%) 20–24.9, n (%) 25–29.9, n (%) ≥30, n (%) Missing, n (%)	26.7 (5.76) 10 474 (2.32) 137 438 (30.5) 190 576 (42.3) 109 585 (24.3) 2597 (0.58)	26.7 (5.76) 10 465 (2.33) 137 321 (30.5) 190 286 (42.3) 109 368 (24.3) 2595 (0.58)	<0.001	– – 0.0859 0.0851 0.1522 0.1980 0.0770	1.05 (1.03–1.07) 1.01 (0.51–1.99) 1. (ref) 1.79 (1.44–2.22) 2.33 (1.86–2.92) –
Smoking status	Never, n (%) Former, n (%) Current, n (%) Missing, n (%)	244 959 (54.4) 155 198 (34.4) 48 010 (10.6) 2503 (0.56)	244 685 (54.4) 154 929 (34.4) 47 923 (10.7) 2498 (0.56)	<0.001	0.1119 0.1733 0.1812 0.1998	1. (ref) 1.55 (1.31–1.83) 1.62 (1.27–2.06) –
Pack years (ever smokers)	Median (IQR) Yes, n (%) No, n (%)	19.5 (23.0) 107 584 (23.9) 341 232 (75.7)	19.5 (23.0) 107 328 (23.8) 340 855 (75.7)	<0.001 <0.001	0.2379 0.1105 0.1079	1.01 (1.01–1.02) 2.16 (1.84–2.53) 1. (ref)
Blood pressure, mmHg	Missing, n (%) Systolic, mean (SD) Diastolic, mean (SD)	1854 (0.41) 138.0 (18.7) 82.3 (10.2)	1852 (0.41) 138.0 (18.7) 82.3 (10.2)	<0.001 <0.001	– – –	1.02 (1.01–1.02) 1.02 (1.01–1.03) 1.00 (0.99–1.01)
Physical activity, METS-h/day	Median (IQR) Missing, n (%)	3.82 (1.67–7.87) 43 671 (9.7)	3.83 (1.67–7.87) 43 592 (9.7)	0.203	–	–
Physical activity, NHS recommended?	Yes, n (%) No, n (%)	235 633 (52.3) 210 147 (46.6)	235 314 (52.3) 209 836 (46.6)	0.263	0.1809 0.1354 0.1480 0.1022	1. (ref) 1.09 (0.94–1.28) – –
Alcohol consumption, units/week	Missing, n (%) Median (IQR)	4890 (1.09) 11.2 (0–24.6)	4885 (1.09) 11.2 (0–24.6)	0.618	–	1.00 (1.00–1.00)
Alcohol, NHS recommended?	Missing, n (%) Yes, <14 units/week, n (%) No, >14 units/week, n (%)	1233 (0.27) 256 718 (57.0) 192 719 (42.8)	1231 (0.27) 256 358 (57.0) 192 446 (42.8)	0.900	0.1622 0.1402 0.1417	1. (ref) 1.01 (0.86–1.18) –
Previous cancer diagnosis	0, n (%) 1, n (%) ≥2, n (%)	409 218 (90.8) 30 385 (6.7) 11 067 (2.5)	408 665 (90.8) 30 326 (6.7) 11 044 (2.5)	0.001	0.1351 0.1942 0.2078	1. (ref) 1.43 (1.10–1.88) 1.54 (1.01–2.34)

IQR, interquartile range; KCa, kidney cancer; METS, metabolic equivalents. \*Tested for difference in means using t-test, difference in medians using Wilcoxon rank-sum test and difference in frequency of cases between categories with Pearson's chi-squared test.

**Table 2** Summary of all validated models.

Models	Type of model	Model description	Individuals included	Age	BMI	Smoking	HBP <sup>a</sup>	Other risk factors <sup>a</sup>	Number of risk factors	All risk factors in primary care records
Chow et al. 1996 [12] (a) (women)	RR	Fully adjusted odds ratios, multiplicative interaction	Women		x	x	x		3	Yes
Chow et al. 1996 [12] (b) (men)	RR	Fully adjusted odds ratios, multiplicative interaction	Men		x	x	x		3	Yes
Colditz et al. 2000 [13] (a) (paper)	Score	Points given for each risk factor based on RR	All		x	x		<b>Family history of kidney cancer</b>	2	Yes
Colditz 2000 [13] (b) (website)	Score	Points given for each risk factor based on RR	All		x	x	x	Height, previous hysterectomy, alcohol consumption, <b>family history of kidney cancer</b>	6	Yes
Colt et al. 2011 [14]	RR	Fully adjusted odds ratios, combined in development	All				x	Ethnicity, anti-hypertensive medication	2	Yes
Hofmann et al. 2015 [15]	RR	Fully adjusted odds ratios, combined in development	All			x	x	Ethnicity, CKD	4	Yes
Hu et al. 2003 [16] (men)	RR	Fully adjusted odds ratios, multiplicative interaction	Men		x	x		Alcohol, fruit, veg and meat consumption	5	No
Hu et al. 2003 [16] (women)	RR	Fully adjusted odds ratios, multiplicative interaction	Women		x	x		Alcohol, education	4	No
Joh et al. 2011 [18] (m1)	RR	Fully adjusted HRs; combined in development	Women		x			Type 2 diabetes	2	Yes
Joh et al. 2011 [18] (m2)	RR	Fully adjusted HRs; combined in development	Women				x	Type 2 diabetes	2	Yes
Joh et al. 2011 [18] (m3)	RR	RR based on number of comorbidities	Women		x		x	Type 2 diabetes	3	Yes
Kadamani et al. 1989 [19] (men)	Score	Linear model	Men		x			Wine consumption, <b>hydrocarbon exposure</b>	2	No
Kadamani et al. 1989 [19] (women)	Score	Linear model	Women		x			Wine consumption	2	No
Kollárová et al. 2012 [20]	RR	Fully adjusted odds ratios, multiplicative interaction	All	x	x	x	x	Sex	5	Yes
Landsman and Graubard 2013 [21] (a)	AR	Logistic regression, unweighted estimator (U)	All	x	x	x	x	Sex, ethnicity, education	7	No
Landsman and Graubard 2013 [21] (b)	AR	Logistic regression, sample pseudo likelihood estimator (SPL)	All	x	x	x	x	Sex, ethnicity, education	7	No
Landsman and Graubard 2013 [21] (c)	AR	Logistic regression, weighted Horvitz-Thompson estimator (W)	All	x	x	x	x	Sex, ethnicity, education	7	No

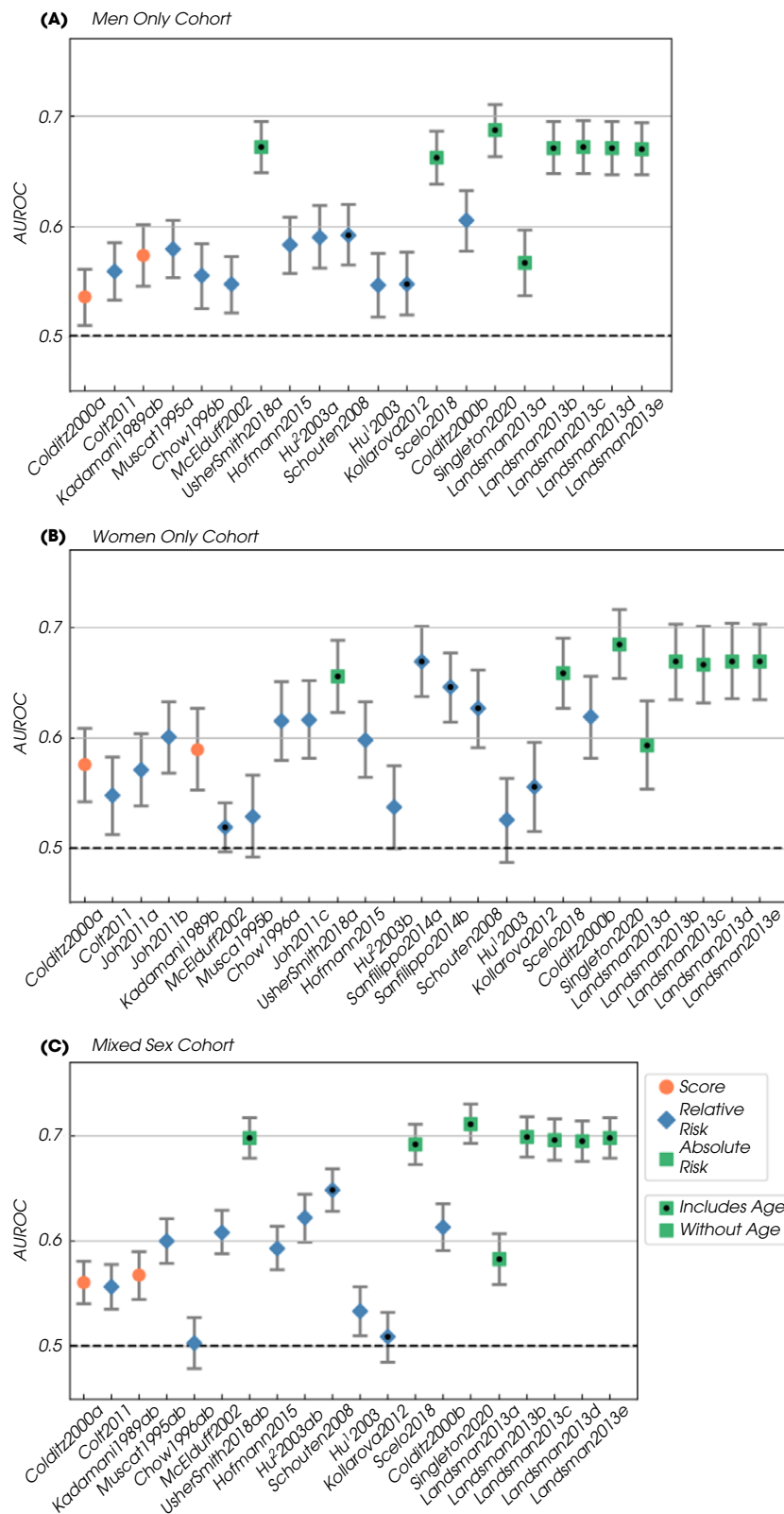


Table 2 (continued)

Models	Type of model	Model description	Individuals included	Age	BMI	Smoking	HBP <sup>†</sup>	Other risk factors*	Number of risk factors	All risk factors in primary care records
Landsman and Graubard 2013 [21] (d)	AR	Logistic regression, re-scaled (of c) weighted estimator (RSW)	All	x	x	x	x	Sex, ethnicity, education	7	No
Landsman and Graubard 2013 [21] (e)	AR	Logistic regression, semiparametric weighted estimator (SPW)	All	x	x	x	x	Sex, ethnicity, education	7	No
McEluff et al. 2002 [22]	RR	Fully adjusted odds ratios, multiplicative interaction	All		x	x		Sex, <b>asbestos exposure</b>	3	Yes
Muscat et al. 1995 [23] (men)	RR	Fully adjusted odds ratios, combined in development	Men		x		x		2	Yes
Muscat et al. 1995 [23] (women)	RR	Fully adjusted odds ratios, combined in development	Women		x		x		2	Yes
Sanfilippo et al. 2014 [24] (systolic)	RR	Fully adjusted odds ratios, multiplicative interaction	Women	x	x	x	x <sup>‡</sup>		4	Yes
Sanfilippo et al. 2014 [24] (diastolic)	RR	Fully adjusted odds ratios, multiplicative interaction	Women	x	x	x	x <sup>‡</sup>		4	Yes
Scelo et al. 2018 [25]	AR	Flexible parametric survival model, 5-year risk	All	x	x	x		Sex, country of residence (UK for all)	5	Yes
Schouten et al. 2008 [26]	RR	Fully adjusted HRs, multiplicative interaction	All	x		x		Sex, alcohol	4	No
Singleton et al. 2020 [7]	AR	Flexible parametric survival model	All	x	x	x	x <sup>‡</sup>	Sex	6	Yes
Usher-Smith et al. 2018 [27] (women)	AR	10-year risk	Women	x	x	x			3	Yes
Usher-Smith et al. 2018 [27] (men)	AR	10-year risk	Men	x	x	x			3	Yes

AR, absolute risk; CKD, Chronic Kidney Disease; RR, relative risk. \*Risk factors in bold are not available for the UKB cohort and were not included in the model validation (note that some are available from online follow-up but were not collected at baseline assessment). <sup>†</sup>High blood pressure (diagnosis prior to baseline). Models that used blood pressure measurement (taken at baseline) indicated by <sup>‡</sup>. Singleton et al. [7] 2020 includes continuous measures of systolic and diastolic blood pressure as well as a history of hypertension.

**Fig. 2** Model discrimination. AUROC value and 95% CIs in (A) whole cohort, (B) men only, and (C) women only. Model type and use of age as a risk factor are indicated on the figure. Model order is determined by number of risk factors used, from left to right, lowest to highest.



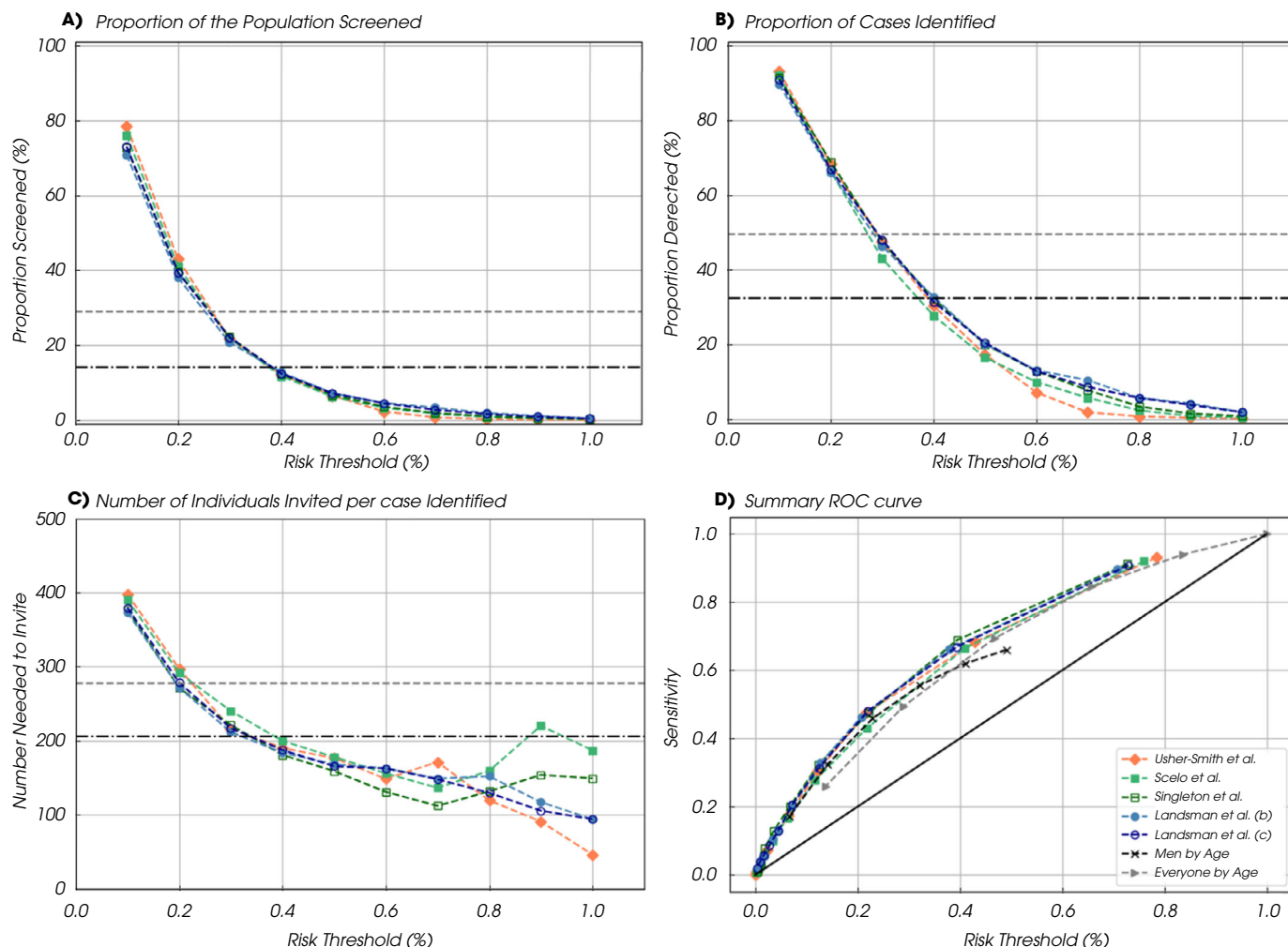


**Table 3** Expected accuracy of a range of screening scenarios, a range of risk thresholds are used to allocate individuals to the high-risk (screening) group.

Screening scenarios	Sensitivity, %	Specificity, %	PPV, %	NPV, %	High-risk individuals, n (%) <sup>*</sup>	Cases detected, n (%) <sup>†</sup>	NNI <sup>‡</sup>
Men aged >60 years	32.4	85.9	0.49	99.8	12600 (14.1)	63 (32.4)	206
Everybody aged >60 years	49.4	71.1	0.36	99.8	28900 (28.9)	105 (49.4)	277
Men aged >65 years	17.0	93.4	0.54	99.8	6600 (6.6)	36 (17.0)	183
Everybody aged >65 years	25.9	86.3	0.40	99.8	13700 (13.7)	55 (25.9)	250
Men aged >60 years. Women aged >65 years	41.5	78.9	0.42	99.8	21200 (21.2)	88 (41.3)	242
Usher-Smith et al. [27]. >0.1%	93.1	21.6	0.25	99.9	78400 (78.4)	197 (93.1)	398
Usher-Smith et al. [27]. >0.3%	47.5	78.4	0.46	99.9	21700 (21.7)	101 (47.5)	216
Usher-Smith et al. [27]. >0.4%	30.3	87.9	0.53	99.8	12200 (12.2)	64 (30.3)	190
Usher-Smith et al. [27]. >0.5%	17.3	93.6	0.56	99.8	6400 (6.4)	36 (17.3)	176
Usher-Smith et al. [27]. >0.7%	1.8	99.3	0.58	99.7	650 (0.65)	4 (1.8)	171
Scelo et al. [25]. >0.1%	92.0	24.0	0.26	99.9	76100 (76.1)	195 (92.0)	391
Scelo et al. [25]. >0.3%	43.0	78.2	0.42	99.8	21800 (21.8)	91 (43.0)	240
Scelo et al. [25]. >0.4%	27.7	88.4	0.50	99.8	11700 (11.7)	59 (27.7)	199
Scelo et al. [25]. >0.5%	16.5	93.8	0.56	99.8	6200 (6.2)	35 (16.5)	178
Scelo et al. [25]. >0.7%	5.7	98.4	0.74	99.7	1600 (1.6)	12 (5.7)	136
Singleton et al. [7]. >0.1%	91.2	27.2	0.27	99.9	72800 (72.8)	193 (91.2)	377
Singleton et al. [7]. >0.3%	47.4	77.9	0.45	99.9	22200 (22.2)	100 (47.4)	221
Singleton et al. [7]. >0.4%	32.2	87.7	0.55	99.8	12300 (12.3)	68 (32.2)	180
Singleton et al. [7]. >0.5%	20.0	93.3	0.63	99.8	6700 (6.7)	42 (20.0)	158
Singleton et al. [7]. >0.7%	7.7	98.2	0.90	99.8	1800 (1.8)	16 (7.7)	111
Landsman and Graubard [21] (b). >0.1%	89.6	29.2	0.26	99.9	70800 (70.8)	190 (89.6)	374
Landsman and Graubard [21] (b). >0.3%	46.3	79.3	0.47	99.8	20700 (20.7)	98 (46.3)	211
Landsman and Graubard [21] (b). >0.4%	32.7	87.3	0.54	99.8	12700 (12.7)	69 (32.7)	183
Landsman and Graubard [21] (b). >0.5%	20.1	92.9	0.60	99.8	7100 (7.1)	45 (20.1)	168
Landsman and Graubard [21] (b). >0.7%	10.4	96.7	0.67	99.7	3200 (3.2)	22 (10.4)	148
Landsman and Graubard [21] (c). >0.1%	90.8	27.0	0.26	99.9	73000 (73.0)	192 (90.8)	380
Landsman and Graubard [21] (c). >0.3%	48.1	78.0	0.46	99.9	22000 (22.0)	102 (48.1)	217
Landsman and Graubard [21] (c). >0.4%	31.4	87.6	0.53	99.8	12400 (12.4)	67 (31.4)	187
Landsman and Graubard [21] (c). >0.5%	20.4	92.9	0.61	99.8	7100 (7.1)	43 (20.4)	165
Landsman and Graubard [21] (c). >0.7%	8.6	97.3	0.68	99.8	2700 (2.7)	18 (8.6)	148

<sup>\*</sup>In a UK population (age range 40–70 years) of 100 000 individuals. <sup>†</sup>Of an expected 212 cases in a UK population of 100 000 individuals aged between 40 and 70 years (over a 6-year follow-up period). <sup>‡</sup>NNI, number of individuals who have to be invited to identify one case.

**Fig. 3** The impact of using the five best performing models in a screening programme is described by (A) the proportion of the population screening, (B) the proportion of cases detected, (C) the number needed to invite and (D) the summary ROC curve, over a range of cut-off thresholds. The relative performance screening all individuals aged >60 years and men aged >60 years are indicated on these graphs.



A lower proportion of cases is detected in women than men by all of the models at this threshold. The Landsman and Graubard [21] (b) model correctly identifies the most cases in women, 11.1% (8/72) compared to 43% (60/139) in men. The screening strategies based on age only  $\geq 60$  years) detect similar proportions of cases in men and women (25.9% and 26.4%); however, the corresponding NNI is much higher for women (249 compared to 189 for men). The Landsman and Graubard [21] and Singleton et al. [7] models classify higher proportions of younger age groups as high risk than other models. In the 55–59-year age group, the Landsman and Graubard [21] (b) model classifies 30.3% of men and 3.9% of women high risk, identifying 50% (14/28) and 21.4% (three of 14) cases, respectively. In comparison, the Scelo et al. [25] model classifies 15.5% of men and 0.2% of women in that group as high risk, accurately identifying 17.9% (5/28) of cases in men and none in women.

## Discussion

We have performed the first comprehensive and systematic validation of risk models ( $n = 30$ ) for kidney cancer in the largest available UK cohort ( $n = 502\,536$ ). In total, 10 models had adequate to good ( $0.61 < \text{AUROC} < 0.72$ ) discrimination in men [7,21,25,27], women [7,21,24,25,27] or the whole cohort [7,21,25,27]. Measures of sex ( $n = 20$ ), smoking ( $n = 22$ ) and BMI ( $n = 29$ ) are included in the majority of models. The highest performing models all included age as a risk factor and several included hypertension. This is consistent with current literature about risk factors for kidney cancer [30,36,37].

The best performing models show discriminative ability comparable to that for models for colorectal cancer [38], breast cancer [39] and melanoma [40] in similar studies. The performance of risk models in kidney cancer based on

phenotypic risk factors is limited by the relatively low population prevalence (0.17, 95% CI 0.09–0.27) in Europe [11]) and relatively weak risk factor associations (e.g. heavy vs never smoker, hazard ratio [HR] 1.83, 95% CI 1.26–2.65 [36]). Models may be improved in the future by the addition of biomarkers or genetic risk factors; however, more research is required to identify suitable candidates [6].

Given the relatively low prevalence of kidney cancer in the general population, age-based screening, such as currently used for colorectal cancer screening in the UK, is unlikely to be implemented. In the public health modelling, the five risk models investigated had slightly better performance than the age- and sex-based strategies in a UK population. The use of any one of these models to identify individuals at higher risk of developing kidney cancer could provide a more nuanced selection process for screening. An ideal screening strategy would provide an efficient method of detecting early stage kidney cancer while minimising the number of unaffected individuals screened. At a threshold of 0.4%, all of the models have an NNI of <200, screening between 11% and 13% of the population and detecting between 27% and 33% of cases. This is comparable to screening all men aged  $\geq 60$  years, where 14.1% of the population are screened and 29.7% of cases are identified (NNI = 206).

The Landsman and Graubard [21] and Singleton et al. [7] models additionally performed particularly well in younger men. At a threshold of 0.4%, the Landsman and Graubard [21] (b) model identified 38% of cases in men aged 45–49 years, although only 7% of this subgroup are classified as high-risk (NNI = 119). Within the context of a screening programme, these models offer the opportunity to identify younger individuals likely to benefit from screening while reducing unnecessary screening amongst low-risk older individuals.

A further important finding is that at a threshold of 0.4%, all of the models classified more men than women as high risk. As a result, using any of the models to determine eligibility for screening would result in a greater proportion of men being invited for screening than women. Given the higher incidence of kidney cancer in men than women (62% of cases are in men in the UK [41]) this may be a reasonable strategy, with men, who are on average at higher risk, being screened and women, who are on average at lower risk, avoiding screening or being screened at older ages. However, all the models also identified a smaller proportion of women who later developed kidney cancer; e.g. the Usher-Smith et al. [27] model detects 44.8% of men who developed kidney cancer but only 2.2% of women, although the NNI for men and women is similar (190 and 192). Using a strategy like this raises ethical considerations about missing more cases in certain subgroups of the population, and equitable distribution of the benefits and harms of screening. Previous

research has shown that selection by a complex risk score is more acceptable to the public than age and sex alone [42].

Recent cost analysis work around kidney cancer screening (using ultrasonography screening) estimated the incremental cost effectiveness ratio to be <£20 000 per quality adjusted life year for men, but not women (when compared to existing practice in the UK) [43]. Using a risk model could result in reduced costs if the population selected screened has a higher prevalence of kidney cancer. However, compared to age- and sex-based strategies, use of these risk models to determine eligibility for screening would require additional data about individuals, increasing the resources required. The Scelo et al. [25] and Usher-Smith et al. [27] models require minimal additional information (BMI and smoking status), often available in primary care records. However, the Landsman and Graubard [21] models require information that is not routinely collected, such as educational level.

## Strengths and Weaknesses

To our knowledge, the present study is the first validation of multiple risk models predicting the development of kidney cancer. Previously only the models by Usher-Smith et al. [27] had been externally validated; our present results are similar to these prior validations [7,27].

A key strength of the present study is the use of the large UKB cohort. In the primary analysis there were 635 cases (408 in men and 227 in women), sufficient for a robust validation [44,45]. Although recruitment for the UKB cohort was designed to be wide reaching, there is evidence of selection bias [46]. The participants differ from the general population in demographics, lifestyle, and health outcomes; the rates of cancer are 11.8% lower than in the UK population [46]. This bias is mitigated in our present analysis by re-scaling predicted risk distributions and standardising the results to reflect the UK population age and sex distribution and kidney cancer incidence. This assumes that cases of kidney cancer in UKB are similar to cases in the UK population but does not address other differences between UKB and the general population. Further, we note that kidney cancer is most common in older individuals (median age at diagnosis is 65 years) and the UKB cohort only covers the age range 40–70 years (46–76 years at study end). Performance of the models in older groups may differ from the results in the present study. Future validations should consider these limitations when selecting an appropriate cohort to test these models.

There is no information in the UKB dataset about histological subtype or stage of cancer at diagnosis. As sub-types of kidney cancer are aetiologically different, it is not clear if the validated models are suitable for predicting the development of all kidney cancers. Many of the models validated were

originally developed for RCC only (Table S1) and most were developed in retrospective cohorts (not specifically designed for screening programmes) [6]. A potential issue with screening is the over-identification of early stage, low-grade cancer. Renal imaging was not carried out for the UKB cohort, so we do not have information about the presence of kidney cancer at baseline and rely on diagnoses in routine care to identify cases. However, as many kidney cancers are currently detected at a late stage (45% Stage III–IV in UK), it is likely that many cases could be detected earlier if a screening programme was implemented. Whether this would translate into a survival benefit is currently unknown and beyond the scope of the present study.

More generally, it is important to consider that variation in performance of the models may also reflect differences in their development populations. Models developed in populations similar to the UKB cohort are likely to perform better in this validation. For example, the Singleton *et al.* [7] model, developed in a pan-European cohort, has good discrimination and calibration in the present study. On the other hand, the Joh *et al.* [18] models, developed in a USA cohort, perform relatively poorly.

In the public health modelling analyses, we assumed that all eligible participants attend screening and that the programme detects all kidney cancers otherwise diagnosed within 6-years. Therefore, the estimates of accuracy are overestimates (and the NNI underestimates) of the real-life implementation of these scenarios. The potential of these models should be considered within the context of a best-practice screening programme and issues including the performance of screening tests, attendance at screening and existing diagnostic pathways should be included in any feasibility studies, as they are in the ongoing Yorkshire Kidney Screening Trial [47]. However, while underlying population prevalence and understanding of the speed with which kidney cancers progress from early to late stage remain uncertain [11], it is not possible to estimate the actual benefits and harms of a kidney cancer screening programme. Randomised controlled trials that compare organised screening to current practise are required understand the role that screening could play in improving health outcomes.

## Conclusions

The present validation study identified five models with good discrimination and calibration in a large UK cohort. These best-performing models are better at identifying individuals who are later diagnosed with kidney cancer than criteria based on age and sex alone, and so could potentially improve the efficiency of a kidney cancer screening programme if used to determine eligibility. However, the benefits of using any of the models over age or sex alone remain small. All of the models additionally struggled to identify women at high risk

of kidney cancer and the number of people needed to invite to identify one individual who is subsequently diagnosed with kidney cancer within 6 years was >100 for most thresholds. Future research is needed to combine these phenotypic models with other biomarkers, including genetic risk factors, to improve performance.

## Acknowledgements

Many thanks to Britt Kilian and Katie Saunders for their advice on analysis. We also thank James Brimicombe for providing data management support.

## Author's Contributions

Hannah Harrison: conceptualisation, methodology, investigation, methodology, validation, software, formal analysis, visualisation, writing – original draft, writing – review and editing, project administration, funding acquisition. Lisa Pennells: methodology, writing – review and editing. Angela Wood: methodology, writing – review and editing. Sabrina H. Rossi: conceptualisation, writing – review and editing. Grant D. Stewart: conceptualisation, writing – review and editing. Simon J. Griffin: conceptualisation, writing – review and editing, supervision, funding acquisition. Juliet A. Usher-Smith: conceptualisation, methodology, writing – review and editing, supervision, funding acquisition, project administration.

## Ethical Approval

The UKB study was approved by the North West Multi-Centre Research Ethics Committee (reference number 06/MRE09/ 65), and at recruitment all participants gave informed written consent to participate in UKB and be followed up, using a signature capture device.

## Conflicts of Interests

Grant D. Stewart has received educational grants from Pfizer, AstraZeneca and Intuitive Surgical; consultancy fees from Pfizer, Merck, EUSA Pharma and CMR Surgical; Travel expenses from Pfizer and Speaker fees from Pfizer. All other authors have no financial disclosures.

## Consent for Publication

Not applicable.

## Funding Information

Hannah Harrison was supported by a National Institute of Health Research (NIHR) Methods Fellowship (RM-SR-2017-09-009) and is now supported by a National Institute of Health Research Development and Skills Enhancement Award (NIHR301182). Sabrina H. Rossi is supported by The Urology Foundation and a CRUK Clinical Research Fellowship. Grant

D. Stewart's work on this topic is funded by Kidney Cancer UK, The Urology Foundation, The Rosetrees Trust, Yorkshire Cancer Research and CRUK and supported by The Mark Foundation for Cancer Research, the Cancer Research UK Cambridge Centre [C9685/A25177] and NIHR Cambridge Biomedical Research Centre. The University of Cambridge has received salary support in respect of Simon J. Griffin from the NHS in the East of England through the Clinical Academic Reserve. Juliet A. Usher-Smith was funded by a CRUK Prevention Fellowship (C55650/A21464) and is now supported by a NIHR Advanced Fellowship (NIHR300861). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

## Data Availability Statement

All data on the risk models are available from the reports or authors of the primary research. All key data fields derived for this study and the underlying STATA code used to generate the main results of the paper will be made available through the UKB access team (access@ukbiobank.ac.uk). No additional data are available.

## References

- Stewart B and Wild CP (eds.) *World Cancer Report*. Lyon: IARC Press; 2014.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424
- Ferlay J, Colombet M, Soerjomataram I et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2019; 144: 1941–53
- Cancer Research UK. Bladder Cancer Statistics. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer> [Accessed: 12/10/2018]
- Vasudev NS, Wilson M, Stewart GD et al. Challenges of early renal cancer detection: symptom patterns and incidental diagnosis rate in a multicentre prospective UK cohort of patients presenting with suspected renal cancer. *BMJ Open* 2020; 10: e035938
- Harrison H, Thompson RE, Lin Z et al. Risk prediction models for kidney cancer: a systematic review. *Eur Urol Focus* 2020 [Online ahead of print]. <https://doi.org/10.1016/j.euf.2020.06.024>
- Singleton RK, Heath AK, Clasen JL et al. Risk prediction for renal cell carcinoma: results from the European Prospective Investigation into Cancer and Nutrition (EPIC) prospective cohort study. *Cancer Epidemiol Biomarkers Prev* 2021; 30: 507–12
- Usher-Smith J, Simmons RK, Rossi SH, Stewart GD. Current evidence on screening for renal cancer. *Nat Rev Urol* 2020; 17: 637–42
- Meisel SF, Side L, Fraser L, Gessler S, Wardle J, Lanceley A. Population-based, risk-stratified genetic testing for ovarian cancer risk: a focus group study. *Public Health Genomics* 2013; 16: 184–91
- Pashayan N, Morris S, Gilbert FJ, Pharoah PD. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: A life-table model. *JAMA Oncol* 2018; 4: 1504–10
- Rossi SH, Hsu R, Blick C et al. Meta-analysis of the prevalence of renal cancer detected by abdominal ultrasonography. *BJS* 2017; 104: 648–59
- Chow WH, McLaughlin JK, Mandel JS, Wacholder S, Niwa S, Fraumeni JF Jr. Obesity and risk of renal cell cancer. *Cancer Epidemiol Biomarkers Prev* 1996; 5: 17–21
- Colditz GA, Atwood KA, Emmons K et al. Harvard report on cancer prevention volume 4: Harvard cancer risk index. Risk index working group, Harvard Center for Cancer Prevention. *Cancer Causes Control* 2000; 11: 477–88
- Colt JS, Schwartz K, Graubard BI et al. Hypertension and risk of renal cell carcinoma among white and black Americans. *Epidemiology* 2011; 22: 797–804
- Hofmann JN, Corley DA, Zhao WK et al. Chronic kidney disease and risk of renal cell carcinoma: differences by race. *Epidemiology* 2015; 26: 59–67
- Hu J, Mao Y, White K, Canadian Cancer Registries Epidemiology Research Group. Overweight and obesity in adults and risk of renal cell carcinoma in Canada. *Soz Präventivmed* 2003; 48: 178–85
- Hu J, Mao Y, White K, Canadian Cancer Registries Epidemiology Research Group. Diet and vitamin or mineral supplements and risk of renal cell carcinoma in Canada. *Cancer Causes Control* 2003; 14: 705–14
- Joh HK, Willett WC, Cho E. Type 2 diabetes and the risk of renal cell cancer in women. *Diabetes Care* 2011; 34: 1552–6
- Kadamani S, Asal NR, Nelson RY. Occupational hydrocarbon exposure and risk of renal cell carcinoma. *Am Journal Ind Med* 1989; 15: 131–41
- Kollárová H, Azeem-Skřejpková K, Tomášková H, Horáková D, Magnusková S, Janout V. Analyzing selected risk factors for the development of kidney cancer. *Cent Eur J Med* 2012; 7: 78–84
- Landsman V, Graubard BI. Efficient analysis of case-control studies with sample weights. *Stat Med* 2013; 32: 347–60
- McElduff P, Attia J, Ewald B, Cockburn J, Heller R. Estimating the contribution of individual risk factors to disease in a person with more than one risk factor. *J Clin Epidemiol* 2002; 55: 588–92
- Muscat JE, Hoffmann D, Wynder EL. The epidemiology of renal cell carcinoma: A second look. *Cancer* 1995; 75: 2552–7
- Sanfilippo KM, McTigue KM, Fidler CJ et al. Hypertension and obesity and the risk of kidney cancer in 2 large cohorts of US men and women. *Hypertension* 2014; 63: 934–41
- Scelo G, Muller DC, Riboli E et al. KIM-1 as a blood-based marker for early detection of kidney cancer: a prospective nested case-control study. *Clin Cancer Res* 2018; 24: 5594–601
- Schouten LJ, van Dijk BA, Oosterwijk E et al. Alcohol consumption and mutations or promoter hypermethylation of the von Hippel-Lindau gene in renal cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 2008; 17: 3543–50
- Usher-Smith JA, Sharp SJ, Luben R, Griffin SJ. Development and validation of lifestyle-based models to predict incidence of the most common potentially preventable cancers. *Cancer Epidemiol Biomarkers Prev* 2018; 13: 13
- Allen N, Sudlow C, Downey P et al. UK Biobank: Current status and what it means for epidemiology. *Health Policy Technol* 2012; 1: 123–6
- Fenton JJ, Weiss NS. Screening computed tomography. *Cancer* 2004; 100: 986–90
- Rossi SH, Klatte T, Usher-Smith J, Stewart GD. Epidemiology and screening for renal cancer. *World J Urol* 2018; 36: 1341–53
- Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd ed., Chapter 5. New York: John Wiley & Sons Inc., 2000.
- Pennells L, Kaptoge S, Wood A et al. Equalization of four cardiovascular risk algorithms after systematic recalibration: individual-participant meta-analysis of 86 prospective studies. *Eur Heart J* 2019; 40: 621–31
- Sun L, Pennells L, Kaptoge S et al. Use of polygenic risk scores and other molecular markers to enhance cardiovascular risk prediction:



- prospective cohort study and modelling analysis. *bioRxiv* 2019. <https://doi.org/10.1101/744565>
- 34 Office of National Statistics (People and Population). Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity>. [Accessed: 02/04/2020]
  - 35 Leiba A, Kark JD, Afek A *et al*. Adolescent obesity and paternal country of origin predict renal cell carcinoma: A cohort study of 1.1 million 16 to 19-year-old males. *J Urol* 2013; 189: 25–9
  - 36 Macleod LC, Hotaling JM, Wright JL *et al*. Risk factors for renal cell carcinoma in the VITAL study. *J Urol* 2013; 190: 1657–61
  - 37 Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. *Eur Urol* 2015; 67: 519–30
  - 38 Usher-Smith JA, Harshfield A, Saunders CL *et al*. External validation of risk prediction models for incident colorectal cancer using UK Biobank. *Br J Cancer* 2018; 118: 750–9
  - 39 Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst* 2010; 102: 680–91
  - 40 Usher-Smith JA, Emery J, Kassianos AP, Walter FM. Risk prediction models for melanoma: a systematic review. *Cancer Epidemiol Biomarkers Prev* 2014; 23: 1450–63
  - 41 Cancer Research UK. Kidney Cancer Statistics. Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/kidney-cancer>. [Accessed: 12/10/2018]
  - 42 Harvey-Kelly LLW, Harrison H, Rossi SH, Griffin SJ, Stewart GD, Usher-Smith JA. Public attitudes towards screening for kidney cancer: an online survey. *BMC Urol* 2020; 20: 170
  - 43 Rossi SH, Klatte T, Usher-Smith JA *et al*. A decision analysis evaluating screening for kidney cancer using focused renal ultrasound. *Eur Urol Focus* 2021; 7: 407–19
  - 44 Moons KG, Wolff RF, Riley RD *et al*. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019; 170: W1–33
  - 45 Wolff RF, Moons KM, Riley RD *et al*. Probst: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019; 170: 51–8
  - 46 Fry A, Littlejohns TJ, Sudlow C *et al*. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol* 2017; 186: 1026–34
  - 47 Yorkshire kidney screening trial. International Standard Randomised Controlled Trial Number (ISRCTN) registry; ISRCTN18055040

Correspondence: Hannah Harrison, Department of Public Health and Primary Care, Forvie Site, Addenbrookes Campus, Cambridge CB2 0SR, UK.

e-mail: hh504@medschl.cam.ac.uk

Abbreviations: AUROC, area under the receiver-operating characteristic curve; BMI, body mass index; CRUK, Cancer Research UK; HR, hazard ratio; NNI, number of individuals who have to be invited to identify one case; NPV, negative predictive value; O:E, observed to expected ratio; PPV, positive predictive value; UKB, UK Biobank.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Model selection process. This flow diagram describes the process used to select models for inclusion in this validation study.

**Figure S2.** Calibration plots comparing the observed and predicted risk of each model in deciles in the UK Biobank cohort.

**Figure S3.** Histograms showing the risk distribution of the five best performing models (a: Usher-Smith, b: Scelo, c: Landsman and Graubard (b), d: Landsman and Graubard (c), e: Singleton) before and after rescaling. The risk is shown separately for men and women.

**Figure S4.** Overview of public health modelling analysis process.

**Table S1.** Summary of the development studies of the models included in the validation.

**Table S2.** The discrimination (AUROC) values for all the validated models in the primary analysis and two sensitivity analyses (open cohort and missing data).

**Table S3.** Calibration measures (observed to expected ratio and Hosmer–Lemeshow test) for all of the models estimating the absolute risk of developing kidney cancer.

**Table S4.** The performance of screening strategies in different sex and age subgroups for a UK population of 100 000 individuals aged 40–70 years. The ability of (a) the five best performing models to identify people at high risk of developing kidney cancer is contrasted to (b) the use of age- and sex-based screening strategies.

**Table S5.** Use of UKB variables.

**Methods S1.** Supplementary Methods Section.